Towards an Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data

Phil Tee
Moogsoft Inc
140 Geary Street, San Francisco, CA 94108
phil@moogsoft.com

George Parisis and Ian Wakeman
School of Engineering and Informatics
University of Sussex
Brighton, UK
{g.parisis, ianw}@sussex.ac.uk

***Abstract*—A key objective of monitoring networks is to identify potential service threatening outages from events within the network before service is interrupted. Identifying causal events, Root Cause Analysis (RCA), is an active area of research, but current approaches are vulnerable to scaling issues with high event rates. Elimination of noisy events that are not causal is key to ensuring the scalability of RCA. In this paper, we introduce vertex-level measures inspired by Graph Entropy and propose their suitability as a categorization metric to identify nodes that are *a priori* of more interest as a source of events.**

We consider a class of measures based on Structural, Chromatic and Von Neumann Entropy. These measures require NP-Hard calculations over the whole graph, an approach which obviously does not scale for large dynamic graphs that characterise modern networks. In this work we identify and justify a local measure of vertex graph entropy which behaves in a similar fashion to global measures of entropy when summed across the whole graph. We show that such measures are correlated with nodes that generate incidents across a network from a real data set.

I. Introduction

An important objective when monitoring a large scale network is detecting failures in critical nodes. This is accomplished by collecting notifications or *events* from the network and analysing these events to determine failed nodes. Events occur at a high rate, and do not always directly indicate a problem. To illustrate, at a typical large enterprise network[1], the event rate is 135 million events a day, generated by just a few hundred 'actionable incidents'.

Identifying which events are the cause of actual outages is called Root Cause Analysis (RCA) [1]. Many algorithms are used to perform RCA [1], but scalability limitations make applying these algorithms to the full event stream impractical. To perform RCA across all events, the flow of events has to be significantly reduced (see for example [2]).

The most common approach to reducing the event rate is the simple act of discarding uninteresting events with a manual filter or exclusion list, a process known as 'blacklisting'. This process is extremely time consuming and error prone. At industrial scale, blacklisting can require thousands of rules; in a fast changing network, such an approach is not practical. A method to automatically eliminate uninteresting events would yield significant savings.

[1]This work is underpinned by the experience at Moogsoft in supplying large scale network management software to many blue chip customers.

In this paper, we introduce a novel technique derived from Graph and Information Theory that determines which events can be treated as noisy, based on the location of their source in the network. The technique relies upon the use of Information Entropy [3], and Graph Entropy [4], [5]. We hypothesise that nodes contributing most to the entropy of a graph are the nodes most likely to generate incidents when events occur. An alternative formulation of the problem is that those nodes contributing most to the connectivity of a graph are most likely to generate incidents when they fail. Graph Entropy is, however, computationally expensive, so we propose alternative formulations that provide similar properties to graph entropy but can be calculated using known global graph properties and information local to the node. We demonstrate that these measures correlate well to the node event pairs that result in incidents.

II. Network structure and other work

After the seminal paper of Barabàsi and Albert [12], there was much work investigating the structure of communication networks, such as by Faloutsos et al [7] and Li et al [8]. The approach primarily focused on datasets generated by discovery protocols such as `traceroute`. This approach was used by Barabási and Albert to assert that communications networks have a power law node degree sequence, possessing the *Scale-Free* property, whereby node degree distributions obey the inverse power distribution law. This was further used to justify the claim that communications networks, like the Internet, are both robust to random attack and vulnerable to targeted attack (the central arguments are outlined in [9], [10], [11], and again in [12]).

The drawbacks of `traceroute` as a discovery protocol are well understood, and outlined clearly in [13] and [14], but essentially arise from the fact that the nature of the `traceroute` tool hides network structure at protocol layers other than IP, and creates many false, high degree nodes. Using more accurate data, built manually from operational change tracking databases of real world networks is a far better way to analyze networks for vulnerability, and includes true connectivity not confined to the IP protocol. We have gained access to a number of datasets from customers of Moogsoft, which number in excess of 200,000 devices and cover many autonomous networks. We can easily dispel the notion of

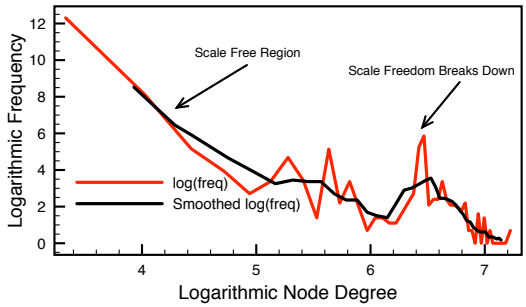

Fig. 1. Scale Freedom Breakdown in a Real Network of 225,239 nodes.

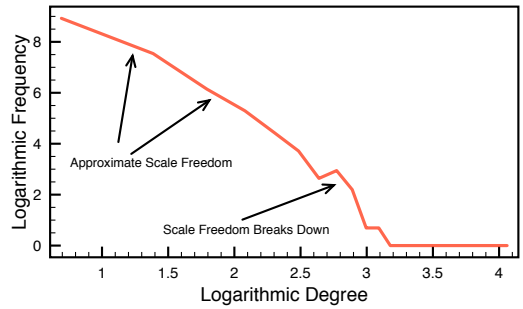

Fig. 2. Scale Freedom Breakdown In the Internet Topology Zoo.

simple power law degree sequences, and hence the generalities implied in [12] and [11], with this dataset, as illustrated in Figure 1 and using the network data published in the Internet Topology Zoo [14] in Figure 2. What is evident from the degree distribution analysis, is that at best the power law is an approximation at low degree, with significant deviations as degree increases. Furthermore for the proprietary dataset the distribution has a notable cluster at high degree values.

Nevertheless, this approach of analyzing communications network using graph invariants, such as node degree, and other related metrics, does indicate that there are methods of identifying nodes which are of more interest from a network vulnerability perspective. The individual contribution of a network node to the overall connectedness of a network, and hence the potential impact of that node failing is an important problem in network management, and the subject of much commercial activity. This has typically been confined to behavioral models of the network (see for example [1], [15]), but these are susceptible to poor scaling behavior on large networks where changes in network topology are frequent. This has particular impact in current networking technologies such as SDN (Software Defined Network), a compelling illustration being [16].

Much focus has been spent in the literature on degree based characterizations of networks from an analysis basis, but it is accepted that degree sequences do not uniquely determine the connectivity properties of a network. Indeed the determination of metrics that allow two networks to be compared for similarity is a much studied and challenging problem in graph theory ([6], [17]). It is the object of this work to establish whether there are other, deeper, node level metrics that can identify the important nodes in a network and yield a useful operational tool to identify operational vulnerabilities of communications networks.

III. Towards Local Measures for Graph Entropy

Historically, entropy has been defined in Graph Theory[2] as a measure of complexity of the global structure of a graph. As a metric it captures many important characteristics, which are of direct interest in a number of applied fields, including the analysis of failure modes of communication networks. In particular, networks with non uniform connectivity will have high values of entropy. Unfortunately the three most well understood measures of entropy involve calculations which have impractical computational complexity, as a graph scales in terms of the number of vertices and edges. What is worse, any change to either the edges or vertices of a graph requires an entirely new computation across the whole graph, and it is extremely difficult to compute the contribution of an individual node to the entropy of the graph. The three variants of Graph Entropy that we shall concern ourselves with are:

- **Chromatic Entropy**: Chromatic entropy is defined by partitioning a graph into sets (or colorings) of disconnected vertices.
- **Körner or Structural Entropy**: The original entropy measure defined on a graph, intended to capture the mutual informational of stable sets.
- **Von Neumann Entropy**: Introduced in analogy to the entropy of quantum systems, this is defined against the eigenvalues of the *Laplacian* matrix associated to a graph.

A valid entropy measure is expected to satisfy the following conditions: *maximality*, *additivity*, *symmetry* and *positivity* [4], [18].

In our treatment we make reference to a number of special graphs, which we define here as:

- K_n **The Complete Graph**: This graph is formed from a set of n vertices, maximally connected.
- S_n **The Star Graph on** n **Vertices**: This graph has one vertex v which is connected to all other vertices, with no other edges in the graph.
- P_n **The Path on** n **Vertices**: This graph is a simple chain of n vertices with no loops, and a start node v_1 and an end node v_n.
- C_n **The Cycle on** n **Vertices**: This graph is a special case of P_n such that $v_1 = v_n$; each node has degree 2.

A central objective of our work is to establish easily computable metrics that measure the contribution of an individual node to the entropy of the whole graph. We will establish that the values, when summed across the whole graph give values consistent with the global measures, and have minimum and maximum values for the same types of special graphs. This analysis establishes the proposed metrics as candidates for

[2]We follow standard graph theory notation for edges and vertices in our presentation.

local vertex entropy measures, and in further work investigate the relationship between the metrics further.

Recent work by Dehmer on Graph Entropy [19],[20] provides a framework that unifies the three global invariants discussed, and provides a pathway to extend these measures in a more computable direction. In particular both Structural and Chromatic entropy rely upon partitions of the vertex set of the graph, which are known *NP-Hard* problems, and, Von Neumann Entropy requires an expensive computation of eigenvalues for the Laplacian Matrix of the graph.

Dehmer defines the concept of a local functional for a vertex, which can be scoped to calculate values for every vertex based upon the local topology of the graph. The degree of locality in the treatment is controlled by using the concept of j-spheres, S_j in the graph, centered at a given vertex. Formally if we denote by $d(v_i, v_j)$ the shortest distance between nodes v_i and v_j, the definition of a j-sphere proceeds as follows:

Definition 1: For a node $v_i \in V$, we define the 'j-Sphere' centered on v_i as:

$$S_j = \{v \in V | d(v_i, v) = j, j \geq 1\} \tag{1}$$

and for convenience later, the related 'j-Edges' E_j as

$$E_j = \{e_{ij} \in E | v_i \in S_j \text{ and } v_j \in S_j\} \tag{2}$$

Using this definition, we then equip each S_j with a positive real-valued function $f_i : S_j \rightarrow R^+$, and further construct a probability functional for each vertex as

$$p_i = \frac{f_i}{\sum_{v_j \in V} f_j} \tag{3}$$

which trivially satisfies $\sum_i p_i = 1$.

The principal direction of Dehmer's proposition is that these functions f_i can be constructed from any structural measure valid and calculable within the 'j-Sphere' of a vertex. In the published work [19],[20], however, these functions are somewhat complex expressions, which introduce global invariants of the graph complicating their computation. We now move on to the important result of this paper, which is the introduction of a variant of Dehmer's approach which uses purely local properties of the neighborhood subgraph of a vertex, and global constants of a graph such as the number of nodes n or the number of edges $|E|$.

A. Local Vertex Entropy Measures

Given the constraint of locality, a number of constructs can be designed which satisfy the probability functional defined in equation (3) up to a normalization constant. It is possible to define the notion of locality using the general concept of j-spheres, but in our treatment we restrict the constructions to a 1-sphere for simplicity of explanation. In the immediate neighborhood of a vertex the available measures are restricted to the degree of the vertex and the presence of cycles in the local subgraph. It is important that the measures that are constructed are bounded in an acceptable way, when summed across the whole graph and satisfy the fundamental properties

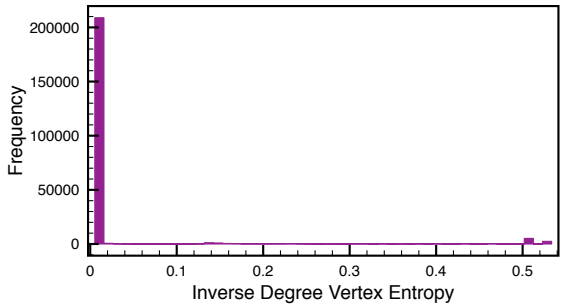

Fig. 3. Frequency Distribution of $VE(v)$.

of an entropy measure: *maximality*, *additivity*, *symmetry* and *positivity* [4], [18].

1) Inverse Degree: The first, and most basic probability functional that we can construct on the 1-sphere of a vertex is its inverse degree. In this case we write the probability at a vertex as:

$$p_i(v_i) = \frac{1}{d_{v_i}} \tag{4}$$

and the corresponding entropy of the vertex $VE(v_i)$, and whole graph $H_{InvDegree}$ as

$$VE(v_i) = \frac{1}{d_{v_i}} \log_2(d_{v_i}),\ H_{InvDegree} = \sum_{i=0}^{i<n} \frac{1}{d_{v_i}} \log_2(d_{v_i}) \tag{5}$$

The first observation is that the sum of inverse degrees does not satisfy the constraint $\sum_i p_i = 1$. However, one can observe that for any given graph G, this probability functional sums to the constant:

$$C = \sum_{i=0}^{i<n} p_i = \frac{\sum_{i=0}^{i<n} \left(\prod_{j \neq i} d_j \right)}{\prod_{i=0}^{i<n} d_i} \tag{6}$$

We note that $p_i = C \times \frac{1}{d_{v_i}}$, and discard the constant as part of the normalization.

We can, however, establish bounds for the value of $H_{InvDegree}$, algebraically. As $p_i < 1$, we can expand (5) to obtain:

$$H_{InvDegree} \approx -\sum_{i=0}^{i<n} \frac{1}{d_{v_i}} \left(1 - \frac{1}{d_{v_i}} \dots \right) \tag{7}$$

Firstly the value is maximized in the case of all degrees being equal and at their maximum. This is satisfied by the complete graph K_n. The minimum requires that the average degree for the graph is at a minimum. This is satisfied by the star graph on n vertices, S_n.

Using the same collection of experimental data used to generate Figure 1, we plot the distribution of values of Inverse Degree Entropy for all nodes in Figure 3. The presence of a large number of edge nodes of degree 1, heavily skews the distribution, but there is a pronounced cluster of nodes at a value of 0.5.

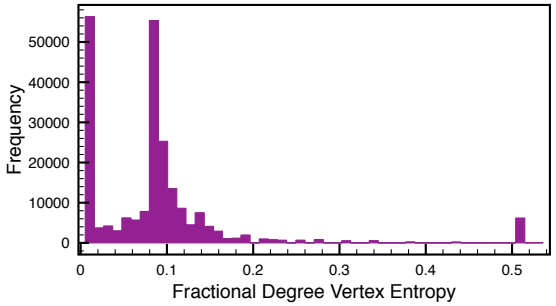

Fig. 4. Frequency Distribution of $VE'(v)$.

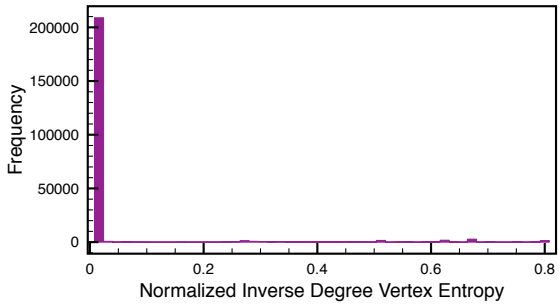

Fig. 5. Frequency Distribution of $NVE(v)$.

2) Fractional Degree Entropy: Inverse degree is unsatisfactory in some regards. Firstly the probability functional is not naturally defined to satisfy the unity sum constraint. Secondly, and more importantly, the degree of a vertex does not capture how '*hub-like*' the node is relative to others. To capture this, we can define an alternative functional, which is based upon the ratio of the vertex degree to the total number of edges in the graph, as follows:

$$p_i(v_i) = \frac{d_{v_i}}{2|E|} \tag{8}$$

Given that $\sum_{v\in V} d(v) = 2|E|$ this functional directly satisfies the unity sum constraint. In a parallel way to equation (5), we define the fractional degree entropy as:

$$VE'(v_i) = \frac{d_{v_i}}{2|E|}\log_2\left(\frac{2|E|}{d_{v_i}}\right) \tag{9}$$

$$H_{FractDegree} = \sum_{i=0}^{i<n} \frac{d_{v_i}}{2|E|}\log_2\left(\frac{2|E|}{d_{v_i}}\right) \tag{10}$$

Following the treatment of Inverse Degree Entropy we establish bounds on this measure by considering the extremal graphs K_n and S_n,P_n. If we expand the logarithmic term in equation (9) we obtain the following higher order term for $H_{FractDegree}$:

$$H_{FractDegree} \approx \sum_{i=0}^{i<n}\left\{\frac{d_i^2}{|E|^2} - \frac{d_i}{|E|}\right\} \tag{11}$$

This is minimized for the graph over n vertices with minimum degree sum, P_n and maximized by K_n.

We plot this value distribution in Figure 4. The distribution of the values is more spread out compared to the Inverse Degree Entropy, but still shows the '*Double Bump*' feature with a cluster centered around a value of 0.1, and a smaller cluster around 0.5. The presence of this '*Double Bump*' in both measures is a necessary but not sufficient condition for these metrics to be useful in highlighting nodes whose impact on connectivity is proportionately higher than others, as both cleanly segregate the nodes into two sets of high and low vertex entropy.

3) Normalized Degree Entropy: There is a considerable practical difference between a star network topology and a fully meshed one, that is between S_n and K_n. In the former, the network is vulnerable to the loss of its central high degree vertex; in the latter, the loss of any one vertex can never create isolated vertices. Both prior measures make no distinction between these two topologies, but there are available metrics measurable at one hop distance that capture this concept. Introduced in [21] and [12] is the concept of the clustering coefficient of a vertex. Utilizing the degree of the vertex i, d_i, it is possible to calculate the fraction of possible edges in the local neighborhood and thereby define the clustering coefficient as:

$$C_i = \frac{2|E_i|}{d_i(d_i+1)} \tag{12}$$

This metric captures how well meshed a node is into its local neighborhood, and therefore serves as an ideal candidate for further refining the vertex measures introduced earlier. In essence, we want to highlight vertices whose clustering coefficient is low, that is their local neighborhood is more similar to S_n locally than K_n. To that end we define the following *Normalized* Vertex Entropies:

Definition 2: We define for a graph $G(V,E)$ the following *Normalized* Inverse Degree Entropy, for both vertex and total graph as follows:

$$NVE(v_i) = \frac{1}{C_i} \times VE(v_i) \tag{13}$$

$$H_{NormInvDegree} = \sum_{i=0}^{i<n} \frac{(d_{v_i}+1)}{2|E_i|}\log_2(d_{v_i}) \tag{14}$$

and the corresponding definition for fractional vertex entropy is defined similarly:

$$NVE'(v_i) = \frac{1}{C_i} \times VE'(v_i) \tag{15}$$

$$H_{NormFractDegree} = \sum_{i=0}^{i<n} \frac{d_{v_i}^2(d_{v_i}+1)}{4|E||E_i|}\log_2\left(\frac{2|E|}{d_{v_i}}\right) \tag{16}$$

Using similar arguments to those used for the non-normalized versions, it is simple to verify that these quantities

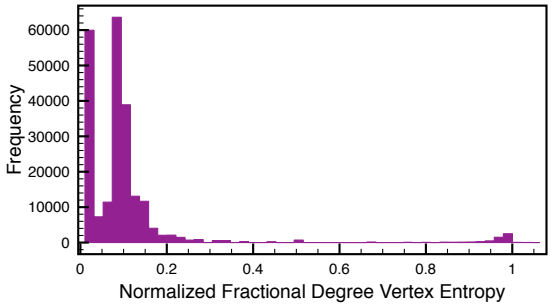

Fig. 6. Frequency Distribution of $NVE'(v)$.

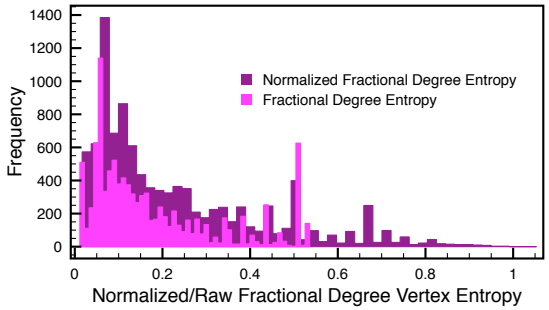

Fig. 7. Frequency Distribution of $NVE'(v)$ & $VE'(v)$ for the Internet Topology Zoo.

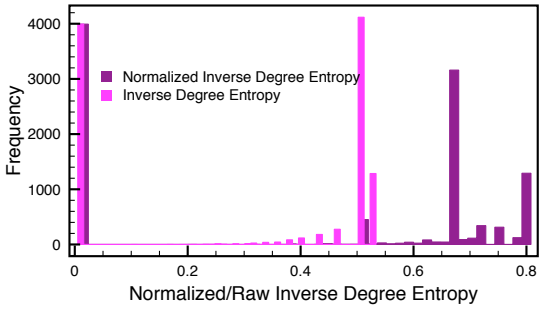

Fig. 8. Frequency Distribution of $NVE(v)$ & $VE(v)$ for the Internet Topology Zoo.

TABLE I
VALUES OF NORMALIZED ENTROPY FOR SPECIAL GRAPHS

	NVE	NVE'
S_n	$\frac{n}{2(n-1)}\log_2(n-1)$	$\frac{1}{2(n-1)}\log_2\{2(n-1)\}+\frac{n}{4}$
K_n	$\frac{n}{n-1}\log_2(n-1)$	$\log_2(n)$
P_n	$\frac{3}{4}(n-2)$	$\frac{1}{n-1}+\frac{3n-4}{2(n-1)}\log_2(n-1)$
C_n	$\frac{3}{4}n$	$\frac{3}{2}\log_2(n)$

are minimized by the graph P_n, and, maximized by the complete graph K_n.

For the same dataset used previously, we plot the distributions of these quantities in Figure 5 and Figure 6. It is interesting to note that both quantities share the same '*Double Bump*' distribution as the non-normalized forms, with a more pronounced separation of the two clusters. We can apply the same analysis to the data in the Internet Topology Zoo [14] and we obtain the distributions in Figure 7 and Figure 8. Although the Internet Topology Zoo is a smaller dataset (19,476 vertices in total), than the proprietary dataset, it still demonstrates a noticeable cluster at high values of both the normalized and raw values of vertex entropy. This '*Double Bump*' style distribution is a necessary, though not sufficient, feature of this metric for it to be useful as a tool in identifying nodes of crucial importance in network monitoring.

To illustrate the bounding values of these normalized quantities for our normalized entropies, summed across our special graphs, we summarize the values in Table I.

From this it is possible to conclude that for NVE, C_n maximizes the value, whereas S_n minimizes it, and for NVE' P_n gives the maximum value and K_n the minimum.

IV. CONCLUSIONS

The main aim of this paper is to introduce computable, node level alternatives to structural entropy measures that are defined across the whole graph. Inspired by the advances made in Barabási's pivotal paper, and suggestions made in the work of Dehmer, we have advanced two computable metrics using structural information available within one hop of a network node. Both of these measures we applied to the proprietary data set, and, to the Internet Topology Zoo data, in both a raw and normalized form. The normalization adjusts the degree based values by the extent to which the local neighborhood of the node is clustered. When these values are applied to the datasets we obtain a distribution which contains two peaks in value, the second peak at higher values of the metric involving far fewer nodes than the first.

The utility of these local measures of entropy *requires* such a distribution if it is to be effective at identifying particular nodes in the networks which introduce vulnerability to the network in terms of overall connectivity. This is more precise than simply selecting the nodes of highest degree, which is central to the scale free argument that a few highly connected nodes, well chosen, represent the bulk of the vulnerability of a network. Nodes with high degree, may be critical to the functioning of the network, but are equally likely to be in a highly meshed and therefore redundant part of the topology. It is the purpose of the normalization of the vertex entropy values to suppress high degree nodes in highly meshed parts of the network, over high degree nodes which are less redundantly wired into the network.

The ultimate test of these values is to examine failure modes in real networks, and, identify if a high value of $NVE(v)$ or $NVE'(v)$ does correlate with those nodes whose failure and removal cause more operational impact on the functioning of the network. For that purpose, we have analyzed the commercial datasets we have access to at Moogsoft and present in Figure 11 an encouraging indication of the utility of one of our measures $NVE'(v)$. We analyzed the distribution

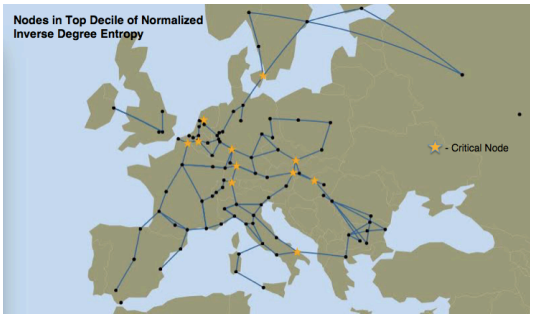

Fig. 9. Critical Nodes in Interoute Network as Identified by $NVE(v)$.

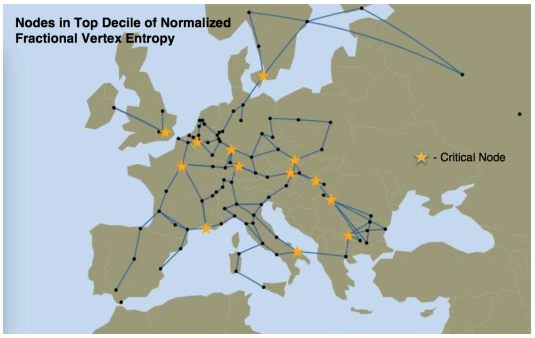

Fig. 10. Critical Nodes in Interoute Network as Identified by $NVE'(v)$.

of over a month of event information from the network and the subset of those events which were escalated by the customers as incidents. It is evident that events distribute around a peak at 0.175 $NVE'(v)$ whereas incidents cluster at a peak of 0.95 $NVE'(v)$.

As further justification of the validity of the approach the detailed nature of the data in the Internet Topology Zoo provides the opportunity to see how the local entropy measures are distributed when calculated against real network topologies. In Figures 9 and 10, we highlight against the *Interoute* topology the top 10% of nodes by value of $NVE(v)$ and $NVE'(v)$ respectively. It is striking to note that in both cases these nodes are indeed at critical points in the graph. For example, the nodes with high values occur at points where their removal would cause the creation of a large disconnected component of the graph, and therefore inflict the highest level of interruption of the operation of the network.

Although the general claims of scale freedom do not fully hold with the real world data we have analyzed in this paper, the motivation to use network structure to identify vulnerable nodes appears promising, and yields two candidates that are locally computable and mirror the behavior of Chromatic and Structural entropy. The justification of studying these values in practical networks has been achieved in theory, and in further work we intend to analyze more real world datasets and extend our entropy measures to include j-spheres for $j > 1$.

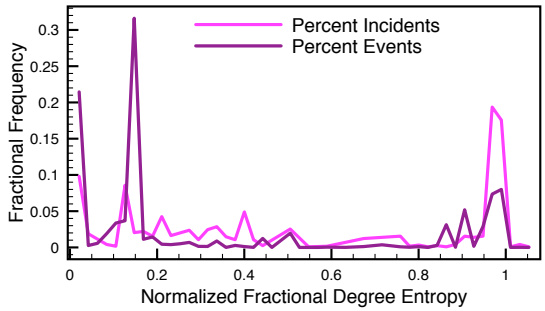

Fig. 11. Distribution of Events and Incidents by $NVE'(v)$ in a real Network of 225,239 Nodes.

References

[1] M. L. Steinder and A. S. Sethi, "A survey of fault localization techniques in computer networks," *Science of Computer Programming*, vol. 53, no. 2, pp. 165–194, nov 2004.

[2] M. Miyazawa and K. Nishimura, "Scalable root cause analysis assisted by classified alarm information model based algorithm," in *Proc. of CNSM*, 2011.

[3] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[4] G. Simonyi, "Graph entropy: a survey," *Combinatorial Optimization*, vol. 20, pp. 399–441, 1995.

[5] J. Körner, "FredmanKomlós bounds and information theory," pp. 560–570, 1986.

[6] B. Bollobàs, *Modern Graph Theory*. Springer-Verlag New York, 1998.

[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *In SIGCOMM*, pp. 251–262, 1999.

[8] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A First-Principles Approach to Understanding the Internet s Router-level Topology," *Acm Sigcomm*, pp. 3–14, 2004.

[9] B. Bollobás and O. Riordan, "Robustness and Vulnerability of Scale-Free Random Graphs," *Internet Mathematics*, vol. 1, no. 1, pp. 1–35, 2004.

[10] B. Bollobás and O. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of Graphs and Networks*. Wiley-VCH, 2006, p. 417.

[11] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–82, 2000.

[12] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Review of Modern Physics*, vol. 74, no. January, 2002.

[13] W. Willinger, D. Alderson, and J. C. Doyle, "Mathematics and the Internet: A Source of Enormous Confusion and Great Potential," *Notices of the AMS*, vol. 56, no. 5, pp. 586–599, 2009.

[14] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 9, pp. 1765–1775, 2011.

[15] S. Kliger, S. Yemini, and Y. Yemini, "A coding approach to event correlation," *... Network Management IV*, 1995.

[16] Y. Tang, E. Al-shaer, and K. Joshi, "Reasoning under Uncertainty for Overlay Fault Diagnosis," *IEEE Transactions on Network Service and Management*, vol. 9, no. 1, pp. 34–47, 2012.

[17] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi, "Graph limits and parameter testing," *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing - STOC '06*, p. 261, 2006.

[18] A. Mowshowitz and V. Mitsou, "Entropy, Orbits, and Spectra of Graphs," *Analysis of Complex Networks: From Biology to Linguistics*, pp. 1–22, 2009.

[19] M. Dehmer and A. Mowshowitz, "A history of graph entropy measures," *Information Sciences*, vol. 181, no. 1, pp. 57–78, 2011.

[20] M. Dehmer, "Information processing in complex networks: Graph entropy and information functionals," *Applied Mathematics and Computation*, vol. 201, no. 1-2, pp. 82–94, 2008.

[21] D. Watts and S. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.